DOCUMENT RESUME

ED 134 951 CS 003 208

AUTHOR Gadway, Charles J.; Wilson, H. A.
TITLE A Handbook of the Mini-Assessment of Functional Literacy--1974 and 1975; Functional Literacy Basic Reading Performance.
INSTITUTION Education Commission of the States, Denver, Colo. National Assessment of Educational Progress.
SPONS AGENCY Office of Education (DHEW), Washington, D.C. Right to Read Program.
PUB DATE 76
NOTE 23p.; See related documents CS 003 211, ED 112 350, ED 112 389

EDRS PRICE MF-$0.83 HC-$1.67 Plus Postage.
DESCRIPTORS Criterion Referenced Tests; Functional Illiteracy; *Functional Reading; *Literacy; *National Competency Tests; National Surveys; Reading Ability; *Reading Achievement; Reading Comprehension; *Reading Research; Reading Tests; Secondary Education
IDENTIFIERS *Mini Assessment of Functional Literacy; *National Assessment of Educational Progress

ABSTRACT

This handbook is designed to give background information on the Mini-Assessment of Functional Literacy, a criterion-referenced test designed to determine the extent of functional literacy among seventeen year olds in America. The five format categories identified for the test were passages; drawings, pictures, signs, etc.; charts, maps, graphs; forms; and reference materials. The five behavior categories selected for the test items were understanding word meanings, gleaning significant facts, comprehending main ideas and organization, drawing inferences, and reading critically. Three standards for comparison are explained: desired level of performance, highest expected level of performance, and minimally adequate performance. The discussion of the methods of describing the data is designed to give the reader of the reports of the Mini-Assessment of Functional Literacy a clearer understanding of the information the data does, or does not, provide. (MKM)

****
* Documents acquired by ERIC include many informal unpublished *
* materials not available from other sources. ERIC makes every effort *
* to obtain the best copy available. Nevertheless, items of marginal *
* reproducibility are often encountered and this affects the quality *
* of the microfiche and hardcopy reproductions ERIC makes available *
* via the ERIC Document Reproduction Service (EDRS). EDRS is not *
* responsible for the quality of the original document. Reproductions *
* supplied by EDRS are the best that can be made from the original. *
****

ED134951

U.S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE
NATIONAL INSTITUTE OF EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

FUNCTIONAL LITERACY

BASIC READING PERFORMANCE

A Handbook of the Mini-Assessment of Functional Literacy -- 1974 and 1975

Prepared for
The National Right to Read Effort
400 Maryland Avenue, SW
Washington, D.C. 20202

In Cooperation With
The National Assessment of Educational Progress
(A Project of the Education Commission of the States)
1860 Lincoln Street, Suite 700
Denver, Colorado 80295

by
Charles J. Gadway, MAFL Coordinator, Analysis and Research
H.A. Wilson, Director, Exercise Development

CS 003 208

NATIONAL ASSESSMENT OF EDUCATIONAL PROGRESS

Roy H. Forbes
Director

1976

The project presented or reported herein was performed pursuant to a Grant from the National Right-to-Read Effort, U.S. Office of Education, Department of Health, Education, and Welfare. However, the opinions expressed herein do not necessarily reflect the position or policy of the U.S. Office of Education, and no official endorsement by the U.S. Office of Education should be inferred.

TABLE OF CONTENTS

ACKNOWLEDGMENTS iv

INTRODUCTION v

CHAPTER 1 Reading and the American Society. 1

CHAPTER 2 Some Standards of Functional Literacy 4

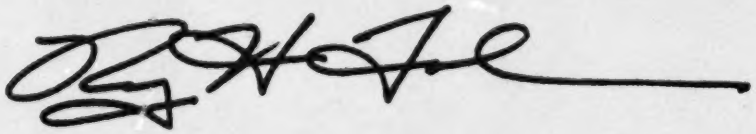
CHAPTER 3 Methods of Describing the Data. 6
Inferring Population Facts From Sample Data 6
Exercise Data 8
Summary Data. 10
Change Data 12
Student Data. 14
Adjusted Data 16

ACKNOWLEDGMENTS

Many people and organizations have made substnatial contributions to the two Mini-Assessments of Functional Literacy (MAFL). Not the least of these are the state and local school systems and the responding students themselves, without whose cooperation these studies would not have been possible. Express gratitude goes to Dr. Ruth Love, former director of the National Right to Read Effort, at whose instigation the MAFL came to be, and to Ms. Helen O'Leary, Right to Read Project Officer for the MAFL. Thanks also go to the other members of the Right to Read staff who contributed their efforts to the success of the MAFL.

The execution of the MAFL was carried out by the following organizations, consultants and National Assessment personnel. To them a sincere thank you.

- For sampling and administration: Research Triangle Institute, Research Triangle Park, N.C.
- For packaging (of exercises) and scoring: Measurement Research Center, Iowa City, Iowa.
- For consultation on the classification of the MAFL exercises; Dr. Alton Raygor, University of Minnesota; Dr. Ruth Hartley, California State College (Sacramento); Dr. Carl Wallen, Arizona State University (Tempe); and Dr. Donald Gallo, Central Connecticut State College.
- For data processing support: Ms. Patrice Gunderson and Mr. John Kalk of National Assessment.
- For statistical analysis consultation and technical review of this report: Dr. Donald Searls and Dr. Robert Larson of National Assessment.



Roy H. Forbes
Project Director

INTRODUCTION

This Handbook of the Mini-Assessment of Functional Literacy is designed as an aid to those who wish to know more about the whys and hows of the Mini-Assessment of Functional Literacy (MAFL) than is given in the two following reports of the results: (1) Brief Summary and Highlights and (2) Statistical/Documentary Report: Summary Volume.

Chapter 1 explains the rationale for the MAFL as derived from the perceived penury of reading skills in America -- especially those skills regarded to be necessary for functioning in daily routines. It also states the rationale and method used for the selection of the reading exercises and their subsequent categorization into sets related to functional reading.

Chapter 2 outlines three standards against which the MAFL results can be evaluated -- each casting functional literacy in a somewhat different light. "Functional literacy" is an abstract term, and without such standards, it can have almost any meaning one may wish to ascribe to it.

Chapter 3 describes and explains the types of data we use in presenting the MAFL results in as clear and popular a manner as possible. Caveats are also given that place limitations on the strengths of the various data and on the interpretations that can be made from them.

CHAPTER 1

READING AND THE AMERICAN SOCIETY

Of all the skills children are taught in school, most Americans would probably agree that reading is the most essential skill for them to acquire if they are to participate even minimally in American society. Even with our greater dependence on the electronic media, the printed word in its many various forms plays a vital role in our lives. Whatever else schools may teach, reading must inevitably be the backbone of the curriculum.

Two hundred years have elapsed since the founding of America -- a nation dedicated to a free and equal education for all. A fitting tribute to our country in this bicentennial era would have been to find that all its citizens are literate, especially those about to graduate from high school either to enter the work-a-day world or to continue their education. Not just functionally literate, but totally literate; these young Americans would be in full and versatile command of reading, writing, speaking and listening.

The bald truth is that not all Americans are literate. The public perception is that far too many are not even functionally literate -- lacking the ability to read, write and converse well enough to function adequately in their everyday lives. If this perception is true, it is a sad commentary on American society that many of its people are illiterate because they may have had inadequate opportunity to learn. It is an even sadder commentary on the American educational system that many of our current generation of youngsters are emerging from that system functionally illiterate.

In recent years, millions of dollars in public funds -- from the federal to local levels -- have been spent on reading programs designed to improve reading skills, particularly of disadvantaged groups. Yet, 200 years after our nation's founding, there is an increasing belief that many young Americans are graduating from (or dropping out of) high school unable to read well enough to function adequately in everyday life. It is not simply that these young people cannot read a great novel in depth; they cannot even comprehend the information in reading materials such as signs, maps, advertisements, forms and reference works so frequently encountered in our day-to-day activities.

Many persons -- both in the educational community and general public -- have "sensed" this tragic situation for some time but

there have been few hard facts to support their belief. In order to provide a body of more concrete evidence as to the extent and nature of this situation, the National Right-to-Read Effort awarded a grant to the Education Commission of the States for the National Assessment of Educational Progress (NAEP) to conduct in two successive years (1974-75) a Mini-Assessment of Functional Literacy (MAFL) of 17-year-old students.

Right to Read and NAEP conducted the MAFL with the principle goal: to determine the degree of functional literacy among 17-year-old students. At first blush, this may appear to be a reasonably simple goal. The accomplishment of this goal, however, implies an evaluation of those reading skills of 17-year-old students considered to be necessary for adequate everyday functioning and, concomitantly, the establishment of one or more standards against which those functional-reading skills can be evaluated.[1] Even prior to this, we must determine which reading skills a person must have in order to be functionally literate. Right to Read has given functional literacy the following theoretical definition.

> A functionally literate person is one who has acquired the essential knowledge and skills in reading, writing and computation required for effective functioning in society, and whose attainment in such skills make it possible for him to develop new aptitudes and to participate actively in the life of his times.

In their day-to-day activities, people encounter many varied types of reading materials such as novels, mystery thrillers, newspapers, magazines, reference works, maps, signs and forms. Depending upon the development of their reading skills, their interest in the material and upon the nature of the reading material itself, people may read "on the surface" or "in depth." That is, a person may be able to just barely understand the meanings of the words; he may be able to unite word meanings to glean isolated facts from the material; or he may relate these facts to recognize the central idea the facts support, draw complex inferences from the facts, or criticize the content. Francis Bacon addressed this concept in his essay, "Of Studies."

> Some books are to be tasted, others to be swallowed, and some few to be chewed and digested; that is, some books are to be read only in parts; others to be read but not curiously; and some few to be read with diligence and attention....

Some types of reading materials, therefore, neither require or merit a deep penetrating study that involves high level reading

---

[1]The standards against which the MAFL reading performances were evaluated are given in Chapter 2.

behaviors or skills. Extrapolating from Bacon's statement, we might say that a fully literate person can, first of all, discriminate between those materials that are best read on the surface and those that require a reading in depth. When he finds materials that need to be "chewed and digested," he is able to do so effectively. A marginally literate person, on the other hand, can at best cope with the shallower types of reading materials. If a person can cope at least with these very basic types of reading materials, he can probably function adequately in everyday life. There is an increasing fear that all too many young Americans are unable to do even this.

The MAFL exercises[2] as a set were selected from a pool of existing NAEP exercises by a panel of reading specialists serving on the Right-to-Read staff who used the dual criteria that they represent those formats of reading materials we frequently encounter in everyday life and that all 17-year-old students should be able to answer them correctly if they are to be able to function adequately. The reading behaviors they elicit are, for the most part, basic.

Additional goals of the MAFL were to determine the degree of functional literacy of the groups of 17-year-old students on the various format aspects of reading materials and on various reading behaviors.[3] The NAEP/MAFL staff and a panel of reading specialists[4] studied the MAFL exercises for both format and the type of reading behaviors they required. Five format categories (passages; drawings, pictures, signs, etc.; charts, maps, graphs; forms; and reference materials) and five behavior categories (understand word meanings, glean significant facts, comprehend main ideas and organization, draw inferences and read critically)[5] were identified. Each exercise part was assigned to one format category and to one behavior category. Data based on these classifications can give us a modicum of insight into the relationship of functional literacy to the format of reading materials and the required reading skills necessary to master them.

---

[2]The MAFL exercises are described briefly in the Brief Summary and Highlights, Appendix B and in the Statistical/Documentary Report: Summary Volume, Appendix B.

[3]A reading behavior can be regarded as a manifestation of a reading skill.

[4]Dr. Alton Raygor, University of Minnesota; Dr. Carl Wallen, Arizona State University (Tempe); Dr. Ruth Hartly, California State University (Sacramento); Dr. Donald Gallo, Central Connecticut State College.

[5]The MAFL format and behavior categories are defined briefly in the Brief Summary and Highlights, Appendix C and more fully in the Statistical/Documentary Report: Summary Volume, Appendix C.

CHAPTER 2

SOME STANDARDS OF FUNCTIONAL LITERACY

The results given in the Mini-Assessment of Functional Literacy (MAFL) reports reflect the functional reading skills of the various groups of 17-year-old students.[1] For the results to be meaningful, they must be evaluated in relation to some standard. In both MAFL reports -- the Brief Summary and Highlights and the Statistical/Documentary Report: Summary Volume[2] -- we present our findings relative to three different standards, each casting the functional-reading skills of 17-year-old students in a somewhat different light.

In both reports, Chapter 2 presents the results relative to a desired performance level (DPL) -- the results we would like to obtain. Since the MAFL exercises were selected on the criterion that all 17-year-old students should be able to answer all the exercises correctly, 100% is the DPL. Therefore, any percentage reported less than 100% represents a shortfall from the DPL standard.

In both reports, Chapter 3 presents the results relative to the highest expected level of performance (HELP) -- the highest results we could reasonably expect to obtain. Here the raw percentages obtained are adjusted according to the percentages obtained by a group of superior readers. After this adjustment, any percentage reported less than 100% represents a shortfall from the HELP standard.[3]

In both reports, Chapter 4 states the results relative to a minimally adequate performance (MAP) standard -- the lowest possible score above which a person would be expected to function at all adequately. Right to Read determined that any 17-year-old student unable to answer at least 75% of the MAFL exercises could reasonably be considered functionally illiterate.

---

[1]These groups are defined briefly in the Brief Summary and Highlights, Appendix A and more fully in the Statistical/Documentary Report: Summary Volume, Appendix A.

[2]The Brief Summary and Highlights gives a briefer form of the results in a less technical presentation; the Statistical/Documentary Report: Summary Volume gives a more thorough treatment of the results from a more technical point of view.

[3]This adjustment is explained in detail in Chapter 3.

The results of the 1974 and 1975 MAFLs[4] and the changes from 1974 to 1975 answer some questions and raise others, but they are not the complete picture. It is impossible for any study, however sophisticated, to measure every aspect of "functional literacy" since there are many definitions of functional literacy.

---

[4]Of the 86 exercises that compose the complete MAFL administered in 1974 and 1975, 64 had been used in National Assessment's 1971 fall reading assessment of 17-year-old students. This subset of exercises, called the truncated MAFL, provides data on the functional-reading skills of 17-year-old students at three time points: 1971, 1974 and 1975.

CHAPTER 3

METHODS OF DESCRIBING THE DATA

Data is like medicine. Both are beneficial when they are used prudently and judiciously for their designated purposes, and attendant cautions are heeded. Likewise, both can be extremely dangerous if they are used indiscriminantly or cautions are ignored.

The following discussion is designed to give the reader of the reports on the Mini-Assessment of Functional Literacy (MAFL) a clearer understanding of the information the data provide and to alert the reader to what the data do not -- indeed cannot -- say.

Inferring Population Facts From Sample Data

From the standpoint of cost in terms of time, effort and money, it was impractical to obtain functional literacy data from all the 17-year-old students in America (the population of interest). Therefore, the National Assessment of Educational Progress (NAEP) obtained data from a sample of the population selected in such a way as to be representative of the population.

Sample data are not perfectly precise. The advantages gained by obtaining data from samples are somewhat lessened by a loss of precision in the descriptions of populations we can give on the basis of that data. Within the limits of measurement error,[1] the data we obtain from a sample precisely describe that particular sample. Even data from a representative sample may not be exactly true for the respective population. When we infer population facts from sample data, we must bear in mind that a very large number of samples -- all representative -- could have been selected from the same population. We would not expect to obtain exactly the same data from all these potential samples. The varia-

---

[1]Measurement error stems from three sources: (1) the measuring instrument -- in this case, the MAFL exercises that may have imperfections such as ambiguity or a built-in tip-off to the correct response; (2) the respondent -- his physical and emotional condition, attitude and motivation; and (3) the measurement situation -- temperature, lighting, pleasantness of surroundings, noise level and the test administrator. These are examples of the three sources of measurement error and are not exhaustive.

tion among the data from these potential samples is called sampling error -- an important concept when we infer population facts from sample data. Most data obtained from the various samples would approximate the population facts quite closely, but the data from some samples could differ by varying degrees.

In order to see more clearly just how this works, consider a classroom of 30 students who have just taken an examination. Twenty-four (80.00%) passed and six (20.00%) failed. These are the population facts. Suppose, however, that we did not know the population facts and wanted to infer them on the basis of data obtained from a randomly selected sample of five students from the classroom. There are 142,506 possible samples of five students that could be selected. If we select our sample in such a way that each student has an equal probability of being selected, our sample is representative of the classroom population.[2] In many of the possible samples of five students that we might select, we would find four (80.00%) who passed the examination and one (20.00%) who failed. In other samples, however, we would find various other percentages of passers and failers. Exhibit 3-1 shows the number having different proportions of passing and failing scores and the percentages of possible samples of five students that could be selected from the 30-student classroom. The variation among the sample percentages of students passing and failing is the sampling error.[3]

The concept of sampling error is important when we consider the percentages of groups of 17-year-old students that correctly answer the MAFL exercises and, likewise, when we compare these percentages from one assessment to the next. A statistic called the standard error can be computed from the one selected sample. It is an estimate of the sampling error -- i.e., the variation among the data for all the potential samples of a given size that could have been selected.

---

[2]A sample can also be representative of the elements if the sample have unequal but known probabilities of being selected. In the latter case, each sample element must be weighted by the reciprocal of its probability of being selected.

[3]It must be noted that this example has been for the purpose of illustration and represents the simplest of all sampling procedures -- a simple sample randomly selected from a small population. The sample NAEP selected for the MAFL was 4,245 17-year-old students from a population in excess of 3 million. The sample was selected in stages from strata (sets of 17-year-old students homogeneous on some trait or characteristic) and clusters (sets of 17-year-old students heterogeneous in the same way as the population). The details of NAEP's sampling procedures are given in The National Assessment Approach to Sampling, which can be obtained by writing to: Ms. Minnie Mitchell, Dissemination Associate, National Assessment of Educational Progress, Suite 700, 1860 Lincoln Street, Denver, Colo. 80295.

EXHIBIT 3-1. Number and Percentages of Possible Samples for Each Combination of Passers and Failers

Passers in Sample		Failers in Sample		Possible Samples	
Number	Percent	Number	Percent	Number	Percent
0	0.00	5	100.00	6	0.0042
1	20.00	4	80.00	360	0.2526
2	40.00	3	60.00	5,520	3,8735
3	60.00	2	40.00	30,360	21.3044
4	80.00	1	20.00	63,756	44.7392
5	100.00	0	0.00	42,504	29.8261
			TOTAL	142,506	100.0000

Exercise Data

While, for the most part, the MAFL reports are concerned with summary data,[4] it is exercise data that is summarized; therefore, it is important to understand exercise-level data first. The basic exercise-level statistic is the p-value (or percentage of success). We also report the standard errors of the p-values.

P-Value (Percentage of Success)

The p-value is the most fundamental NAEP statistic. Most simply stated, a p-value is an estimate of the percentage of a group who would have made a correct response to an exercise had it been administered to every member of a group. It is alternatively referred to in the MAFL Brief Summary and Highlights report as the percentage of success.

Although the NAEP-MAFL samples of 17-year-old students were selected to ensure that they are representative of their respective populations,[5] each student in a population did not have exactly the same probability of being selected for the sample as was the case in our classroom example; that is, each respondent in a MAFL sample did not represent the same number of 17-year-old students in the population. Therefore, each respondent in a sample was assigned a weight that is the reciprocal of the probability of his being selected for the sample.[6] The larger the respondent's

---

[4]We report exercise data only for a few selected exercises that are unique in some manner.

[5]The populations of 17-year-old students and their subclassification are defined briefly in the MAFL Brief Summary and Highlights, Appendix A and more thoroughly in the MAFL Statistical/Documentary Report: Summary Volume, Appendix A.

[6]Weighting is necessary only when the members of a population have different but known probabilities of being included in a sample. See also footnote 2.

selection probability, the smaller his weight since he represents fewer 17-year-old students in the population. A p-value (percentage of success) is the ratio of the sum of weights of those responding correctly to the sum of weights of all the respondents in the sample.

For each p-value that we report, we also report its standard error. This statistic can be used to construct one or more ranges of p-values within which we can express some degree of confidence that the population p-value occurs. Such ranges of p-values are called confidence intervals.[7] Two rules should be noted:

1. The smaller the standard error, the smaller the confidence interval will be for any specific degree of confidence that we might desire.

2. For any specific standard error, the higher the degree of confidence used, the larger the confidence interval will be.

Like data, many events in our day-to-day lives lack pinpoint precision; and we frequently encounter "expectancy ranges" of various types applied to events. Such expectancy ranges lack the statistical ramifications of confidence intervals and are formed in a much less precise way. They are, however, based on a person's past experiences in the event in question. Consider this example. You go to a store to purchase a particular product. The salesperson informs you that the product is out of stock but that he will order it for you. He estimates that it will take a week, give or take two days, for the product to arrive. The salesperson knows from past experiences that "most often" the product arrives within the five-to-nine-day interval and "only rarely" arrives sooner than five days or later than nine days.

Limitations of Exercise Data

As we have shown in the foregoing sections, sample data are not perfect; but if used prudently, they can provide a wealth of information. Within the limits of measurement error and sampling error, the data presented in the MAFL reports accurately describe the functional-reading achievements of the various groups of 17-year-old students.

When the data show that one group has achieved either above or below another, one must exercise caution in attributing causation to this difference. Many factors may affect the performance

---

[7]The reader is referred to any book on basic statistics for aid in constructing confidence intervals.

of 17-year-old students in a given group on the MAFL exercises. Consider, for example, a hypothetical group whose functional-reading achievement is above all other groups. Most members of the group may attend schools that have high-quality faculties, excellent reading programs and a large library; have parents who place a high value on education and encourage reading; and have available at home a wide variety of reading materials. All these factors could contribute to the group's high level of achievement while membership in the group itself may contribute very little or nothing.

The name of a group is merely a categorical label, and the traits attributable solely to that label must not be construed as necessarily being the cause or even as being a cause for the group's comparatively high or low achievement.

Summary Data

In summarizing the MAFL results, we wish to describe each group's functional-reading achievements over the entire set of MAFL exercises and over the subsets of exercises classified under the five formats and five behaviors defined in Appendix C of both the MAFL Brief Summary and Highlights and the MAFL Statistical/Documentary Report: Summary Volume. To accomplish this, we need a number that best represents the set of data being summarized. One such number is the mean. It is an average of the set of numbers being summarized that takes into account the value of each number being summarized. It is a statistic upon which addition and subtraction can be performed. It is amenable to measuring change over time since the mean of the differences between the p-values at two time points is equal to the difference between the means of the p-values at those time points.

Mean P-Value (Mean Percentage of Success)

The mean p-value (mean percentage of success) is the sum of the p-values for the set of exercises being summarized divided by the number of exercises in the set. We report mean p-values for all 17-year-old students and for each of the subclassifications of 17-year-old students (see footnote 5) on all 86 MAFL exercise parts and on each of the five format and five behavior classifications.[8,9] We also report mean p-values for all groups of 17-year-

---

[8]The mean p-values for the format and behavior classifications are tabulated only in the MAFL Statistical/Documentary Report: Summary Volume. Those that are noteworthy are reported and discussed in the MAFL Brief Summary and Highlights.

[9]The format and behavior classifications of the MAFL exercises are defined briefly in the MAFL Brief Summary and Highlights, Appendix C and more thoroughly in the MAFL Statistical/Documentary Report: Summary Volume, Appendix C.

old students on the 64 exercise parts in the truncated MAFL but not for the format and behavior classifications within this set.[10]

Standard Deviation[11]

While the mean p-value is a good general indicator of a group's achievement on some functional-reading skill, it tells us nothing about the spread of the individual p-values around the mean p-value. Two groups could have identical mean p-values on some functional-reading skill. This gives the impression that they have the same achievement on the skill. In one sense, this is true; but if the individual p-values of one group cluster tightly about the mean, and those of the other group deviate widely from the mean; the two groups obviously are not alike. The standard deviation is used to indicate the spread of individual p-values around the mean. For example, consider the following sets of percentages.

(A) 10, 20, 30, 40, 50, 60, 70, 80, 90; and
(B) 40, 40, 40, 50, 50, 50, 60, 60, 60.

Both A and B have a mean equal to 50%, but the standard deviation of A is equal to 27.39, while that of B is 8.66. The larger the standard deviation, the greater is the spread of the data. For each set of p-values summarized, we report the standard deviation along with the mean p-value in the MAFL Statistical/Documentary Report: Summary Volume.

Limitations of Summary Data

All the limitations of exercise data given previously apply to summary data as well. In addition, certain special limitations apply to summary data.

---

[10] Of the 86 exercises that comprise the complete MAFL administered in 1974 and 1975, 64 had been used in National Assessment's 1971 fall reading assessment of 17-year-old students. This subset of exercises, called the truncated MAFL, provides data on the functional-reading skills of 17-year-old students at three time points. Within the truncated MAFL, most formats and behaviors contain too few exercises to be summarized meaningfully.

[11] The standard deviation should not be confused with the standard error discussed previously. The standard deviation indicates the amount of spread of the p-values in a set around the mean p-value of that set. The standard error is a special form of standard deviation that is an estimate of the spread that would occur among the mean p-values of all the potential samples (of a given size) that could be selected from a population (see the discussion of sampling error in this chapter).

Whenever data are summarized, some information is lost. The mean p-value is a number that describes a group's overall level of performance on some functional-reading skill, and the standard deviation indicates the spread of the individual exercise p-values. These numbers do not tell us, however, on which exercises a group performed quite differently than we would have expected on the basis of the group's overall performance. If the mean is "the rule," remember that there are exceptions to every rule. This becomes more important when the standard deviation is quite large.

Another problem that must be taken into account when considering summary data is that of exercise sampling. We have stated previously that our samples of 17-year-old students are representative of the populations from which they were selected. We could, therefore, infer population facts from sample data. We would like to be able to say this about our samples of exercises, but we simply cannot. In the first place, it was not possible to identify populations of functional-literacy exercises -- all the possible reading exercises that measure functional literacy or the various aspects of it (i.e., formats and behaviors) -- and then select representative samples. The MAFL exercises were selected by a panel of reading specialists who, by consensus, agreed that they measured functional literacy. Even with this face validity, the question remains: Is the selected sample of functional-literacy exercises representative of all possible functional-literacy exercises? There is no objective answer to this question. The MAFL exercises were categorized by the consensus of another panel of reading specialists into the formats and behaviors. Since these samples are smaller -- and in some cases very small -- the question of representativeness is even more serious. Caution must be exercised, therefore, in generalizing summarized data -- whether on the entire set of MAFL exercises or one of the formats or behaviors -- to the respective populations of exercises, and the smaller the number of exercises being summarized, the greater the caution we must use.

Change Data

We report two lines of changes over time in the functional-reading achievements of 17-year-old students. One pertains to the complete set of 86 MAFL exercises administered first in 1974 and repeated in 1974 and to the format and behavior classifications within this set. The second pertains to the truncated subset of 64 MAFL exercises that had been first administered to 17-year-old students in NAEP's 1971 assessment of reading.

The Complete MAFL

A change in a functional-reading skill from the 1974 MAFL to the 1975 MAFL at the summary level is computed by subtracting the mean p-value (mean percentage of success) of the 1974 MAFL from

the mean p-value of the 1975 MAFL. If the resulting change has a positive value, a gain in achievement is indicated; if the resulting change has a negative value, a loss in achievement is indicated. We compute for each change in mean p-values a number called the critical ratio by dividing the change by the standard error of the change. From these critical ratios, by using a table of the normal distribution, we determine the probability that each observed sample change could have occurred due to chance (sampling error).[12] We report these exact probabilities and the standard error for each change in both the MAFL Brief Summary and Highlights and Statistical/Documentary Report: Summary Volume. Right-to-Read has decided that for its purposes, a change must have a probability no larger than 0.050 that it might have occurred due to chance.[13] Only changes meeting this criterion are given in the text of the MAFL Brief Summary and Highlights. These changes are marked with an asterisk (*) in the tabular presentations in both the MAFL Brief Summary and Highlights and the MAFL Statistical/Documentary Report: Summary Volume.

The Truncated MAFL

For the 64 exercises in the truncated MAFL, we report mean p-values (or mean percentages of success) at three time points: 1971, 1974 and 1975. We report the changes from 1971 to 1974, 1974 to 1975, and the net change from 1971 to 1975 in the same manner as the 1974-75 changes for the complete MAFL discussed above.

Limitations of the Change Data

All the limitations of exercise data and summary data given previously apply to change data as well. In addition, certain other cautions must be observed when considering change data.

When a given subject area, such as MAFL, has been assessed at two or three points in time, it only begins to be possible to determine whether that subject area is progressing or regressing. Attaching too much importance to an observed gain or loss in achievement between any two successive points in time can be misleading in terms of a group's trend in achievement over long periods of time -- even if the observed gain or loss is statistically significant.[14] On the other hand, a series of nonsignificant

---

[12]See earlier discussion on sampling error.

[13]A test of change such as this only provides evidence regarding the direction of the population change. It does not imply that the magnitude of the population change is the same as the magnitude of the sample change. It may be larger or smaller.

[14]There is no greater than a 0.050 probability that the change occurred due to chance.

changes in the same direction over a number of assessments may provide more information about a group's achievement trends than a single significant change.

It is true of most measures of change -- and certainly true of changes in educational achievement -- that they cannot be expected to be uniform or even always to be in the same direction over several measures. The important aspect of change in general is the trend of improvement or decline over several assessments. The important aspect of a change between any two adjacent assessments is not so much whether that particular change is statistically significant but whether it represents a continuation of improvement or declination or departs from it.

An improvement or decline that departs from an established trend may be spurious due to sampling variability, or it may augur the establishment of a new trend. This can be determined only after additional assessments. If a sharp departure from a trend (say, an improvement) is spurious, it will quite likely be followed by a sharp reversal (in this case, a decline) on the succeeding assessment indicating a probable return to the established trend.

Student Data

Another, somewhat different, way to examine the functional-reading skills of 17-year-old students is based on student or respondent scores (s-values) on the package he took.[15] While the concern of the MAFL is not with the achievement of individual respondents per se, student scores are useful in making statements like "Seventy-two percent of 17-year-old students could correctly answer at least 85% of the MAFL exercises."

S-Values (Student Scores)

Since all the MAFL exercises have unit weights, an s-value is simply the number of exercises answered correctly by a given student. Since each student answers exercises in only one of the two MAFL packages, observed s-values are package specific. If the two MAFL packages are A and B, we have observed s-values, S_A and S_B for the 17-year-old students who answered packages A and B, respectively. Of greater interest than how well a 17-year-old student performed on the exercises in one package, is how well he would have performed on all the MAFL exercises had he taken them. A method has been developed for estimating each student's score on the package he did not take from his score on the package he

[15]To preserve a student's anonymity, his name is never entered on a package; and individual scores are never reported to anyone.

took.[16] Thus, for each student taking package A, we have his score on package A (S_A) and an estimate of what his score would have been on package B ($\hat{S}_B$). Likewise, for each student taking package B, we have his score on package B (S_B) and an estimate of what his score would have been on package A ($\hat{S}_A$). If, for each student taking package A, we add his s-values S_A and $\hat{S}_B$, we have an estimate of the number of all MAFL exercises he would have answered correctly ($S_{M/A}$). And if, for each student taking package B, we add his s-values S_B and $\hat{S}_A$, we have an estimate of the number of all MAFL exercises he would have answered correctly ($\hat{S}_{M/B}$). Next, if we convert $\hat{S}_{M/A}$ and $\hat{S}_{M/B}$ to percentages ($\hat{S}_{M/A}$ and $\hat{S}_{M/B}$, respectively), these percentages can be averaged to obtain a generalized estimate of the percentage of MAFL exercises ($\hat{S}'_M$) a 17-year-old student would have obtained had he taken all MAFL exercises where:

$$\hat{S}'_M = (\hat{S}'_{M-A} + \hat{S}'_{M-B})/2.$$

The major premise of this transformation is that a 17-year-old student will maintain the same rank-order position of achievement on different sets of exercises measuring the same (or nearly the same) construct.

How Many Perform How Well?

The major purpose of the s-value is to report the percentages of 17-year-old students in the nation and in each of the subclassification groups who can correctly answer various percentages of the MAFL exercises. In the MAFL Brief Summary and Highlights, we report the percentages who correctly answered at least 95, 85, 75 and 65% of the MAFL exercises. In the Statistical/Documentary Report: Summary Volume, we report the percentages who correctly answered at least 95, 90, 85, 80, 75, 70, 65, 60, 55, 50 and 25% of the MAFL exercises is the minimally acceptable performance level standard for functional literacy designated by Right-to-Read and those 17-year-old students falling below this standard can be reasonably classified as functionally illiterate.

These percentages are given for both 1974 and 1975 and the changes from 1974 to 1975, the standard errors of the changes, the critical ratios,[17] and the probabilities that the changes are due to random error. Changes for which the probability is no greater than 0.050 are marked with an asterisk (*).

---

[16]A separate paper on this method can be obtained by writing to: Dr. Donald T. Searls, National Assessment of Educational Progress, Suite 700, 1860 Lincoln Street, Denver, Colo. 80295.

[17]Critical ratios are reported in the MAFL Statistical/Documentary Report: Summary Volume only.

Limitations of Student Data

In general, the limitations that apply to exercise data, summary data and change data also apply to student data.

Adjusted Data

It is often the case that data can be misunderstood or misinterpreted since there is no standard against which it can be evaluated. In the MAFL, we have been fortunate in that 100% is the desired performance level so that the MAFL data can be evaluated against this standard. While it may be desired or ideal that 100% of 17-year-old students be able to answer all the MAFL exercises correctly, the NAEP-MAFL staff hypothesized that it may not be the most realistic for determining their degree of functional literacy. If, for example, a group's mean p-value (percentage of success) is taken as an index of that group's degree of functional literacy, it would logically follow that the group's mean q-value (100% minus mean p-value) is an index of the group's degree of functional illiteracy. That is, any p-value or mean p-value less than 100% would indicate some degree of functional illiteracy. We believed that the achievement level on each MAFL exercise by a group of known superior readers provides a more realistic standard of functional literacy. We assumed that these superior readers would be functionally literate to the highest expectable degree and that their MAFL achievements (p-values) would represent the highest expected levels of performance (HELPs) for the MAFL exercises.

We arbitrarily define a superior reader as a 17-year-old student who had attained at least the 90th percentile on the College Entrance Examination Board reading test or an equivalent standardized reading test.[18] We located 100 superior readers in the Denver metropolitan area[19] and administered both MAFL packages to them. The percentage of superior readers that responded correctly to each exercise was considered to be the HELP for that exercise. Raw p-values (percentages of success) have been adjusted to the HELP standard by converting each to a percentage of the HELP for that exercise according to the formula:

$$\tilde{P}_{ij} = (P_{ij}/P_{sj})\ 100, \text{ where;}$$

$\tilde{P}_{ij}$ = adjusted p-value for group i on exercise j;

---

[18]The reading tests actually used were: The Nelson-Denny Reading Test; Stanford Achievement High School Battery; and Comprehensive Test of Basic Skills, California Test Bureau.

[19]Since superior readers are homogeneous on the trait of reading ability, we deemed 100 to be an adequate, it not ideal, sample size, and that it was unnecessary to classify them by region, sex, race, parental education or type of community.

P_{ij} = raw p-value for group i on exercise j;

P_{sj} = p-value for superior readers on exercise j.

For example, if a group achieved a p-value of 70% on a given exercise and the superior readers achieved a p-value of 85% on that same exercise, 85% would be the HELP for that exercise and the group's adjusted p-value would be 100 (70%/85%) or 82.35%.

To summarize the performance of a group on all the MAFL exercises or one of the subsets relative to the HELP, we compute the mean adjusted p-value[20] in the same manner as the raw p-value.

[20]The mean adjusted p-value is the mean of the adjusted p-values and numerically differs slightly from the adjusted mean p-value.